ABSTRACT

A spatial knowledge base approach for the automatic extraction of metadata **116** from electronic documents **100**. The electronic document **100** is converted to a substantially

5   format invariant data file **104** by an intermediate language conversion element **102**. Spatial layout facts **108** are extracted and combined with spatial layout rules **114** from a knowledge engineer **112** in a spatial metadata-reasoning element **110** to provide the metadata **116**. The invention is based on mimicking the visual and spatial knowledge that humans make use of when reading a document.

10